

# Accuracy of Facial Forensic Examiners and Super-Recognizers

**Dr. P. Jonathon Phillips**

*National Institute of Standards and Technology*



# Human Accuracy on Hard Cases

## biometric TECHNOLOGY TODAY

ISSN 0969-4765 April 2016

www.biometrics-today.com

FEATURE

## Humans vs machines: the future of facial recognition

Tim Ring, journalist



Tim Ring



London Metropolitan Police Commissioner Sir Bernard Hogan-Howe (left) presents the force's Staff of the Year award to 'super recogniser' Detention Officer Idris Bada.

Courtesy T. Ring, *Biometric Technology Today*, 2016

NIST

# Outline

- Range of human accuracy
- Perceptual accuracy of facial forensic examiners
- Accuracy of super-recognizers
- Measure accuracy of facial forensic examiners in their milieu

# Facial Examiner vs Super-recognizer

- Two different scenarios
- Super-recognizers (The Met)
  - Familiar face recognition
  - From memory
  - Learned faces on their “beat”
- Facial Examiners
  - Unfamiliar face recognition
  - Compare faces side-by-side
  - Trained
  - Tools and methods

# Glasgow Face Matching Test

**Question: What is the range of human accuracy?**

Match



Nonmatch



**Same or different?**

Burton, White & McNeill (2010). *Behavior Research Methods*, 42, 286-291.

# Range of Accuracy on GFMT

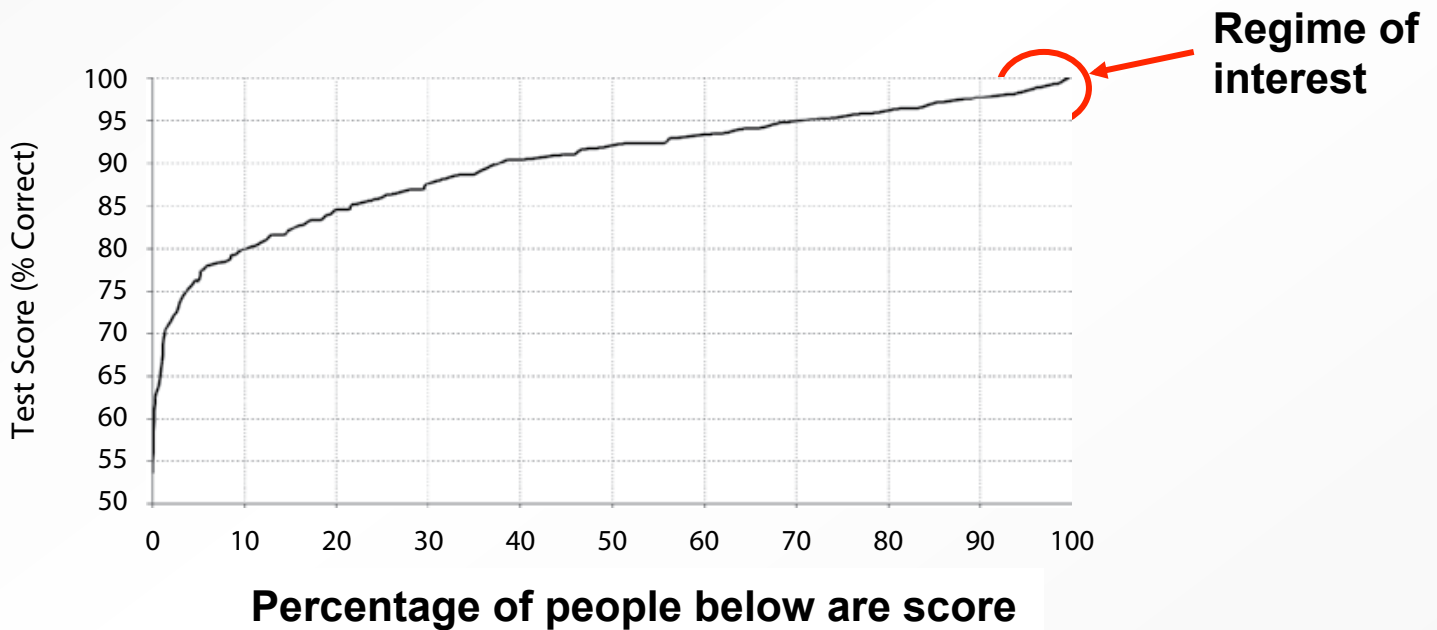
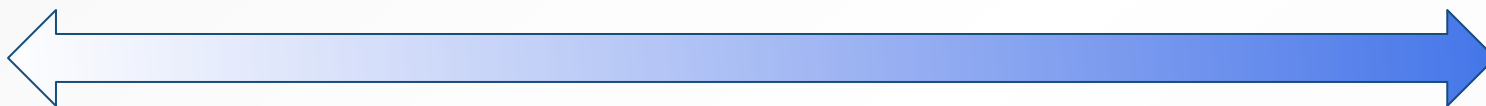


Figure 2. Cumulative frequency of accuracies for the Glasgow Face Matching Test.

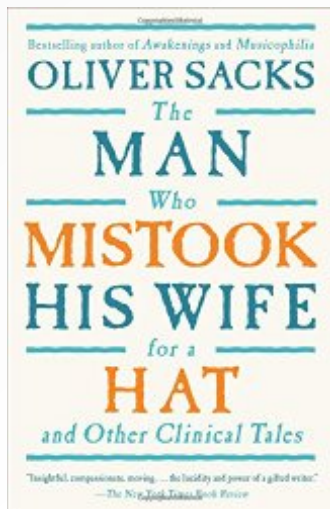


# Two Dimensions of Recognition

Perceptual



Low aptitude

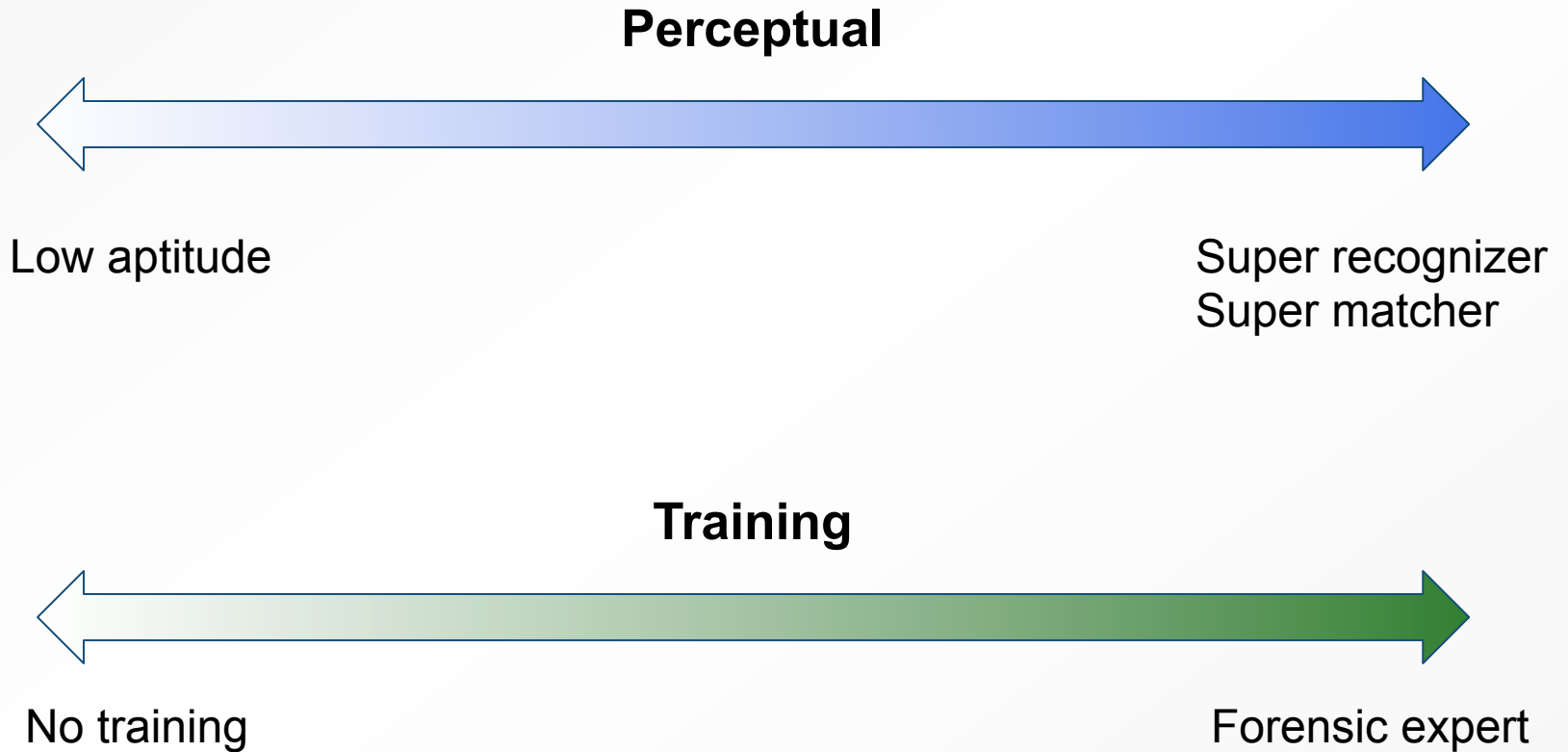


Super recognizer  
Super matcher



London Metropolitan Police Commissioner Sir Bernard Hogan-Howe (left) presents the force's Staff of the Year award to 'super recogniser' Detention Officer Idris Bada.

# Two Dimensions of Recognition





# Proceedings of Royal Society: B

## PROCEEDINGS B

[rspb.royalsocietypublishing.org](http://rspb.royalsocietypublishing.org)



CrossMark  
click for updates

### Research

**Cite this article:** White D, Phillips PJ, Hahn CA, Hill M, O'Toole AJ. 2015 Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B* **282**: 20151292. <http://dx.doi.org/10.1098/rspb.2015.1292>

## Perceptual expertise in forensic facial image comparison

---

David White<sup>1</sup>, P. Jonathon Phillips<sup>2</sup>, Carina A. Hahn<sup>3</sup>, Matthew Hill<sup>3</sup> and Alice J. O'Toole<sup>3</sup>

---

<sup>1</sup>School of Psychology, The University of New South Wales, Sydney, New South Wales 2052, Australia

<sup>2</sup>National Institute of Standards and Technology, 100 Bureau Drive, MS 8940, Gaithersburg, MD 20899, USA

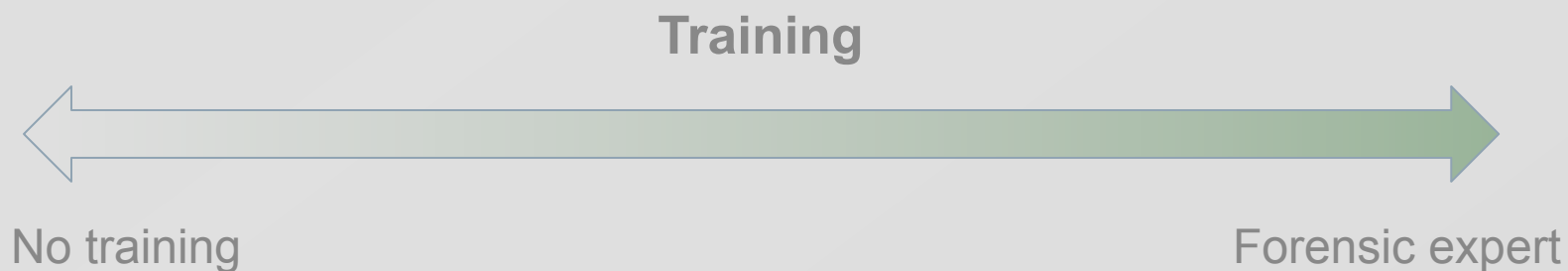
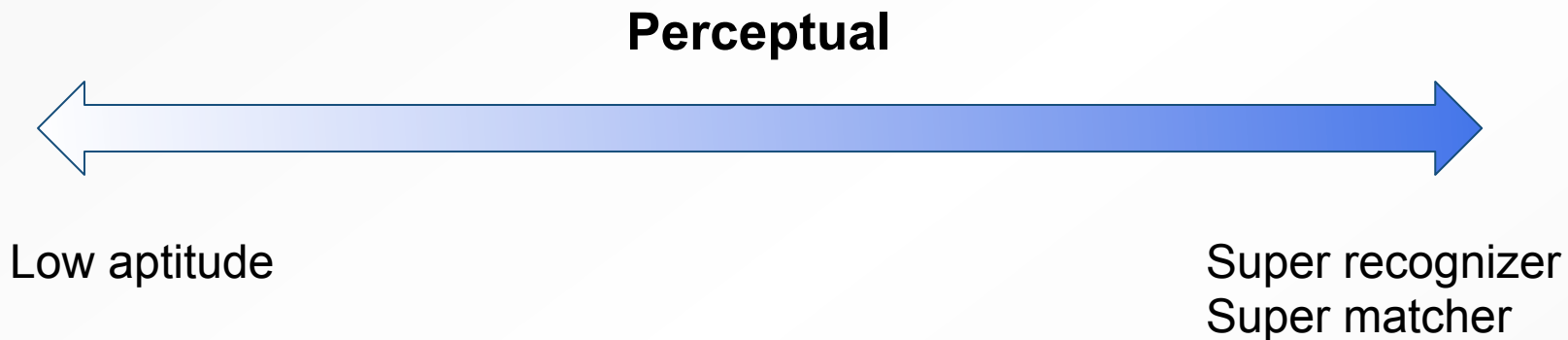
<sup>3</sup>The University of Texas at Dallas, Richardson, TX 75080, USA

Forensic facial identification examiners are required to match the identity of faces in images that vary substantially, owing to changes in viewing conditions and in a person's appearance. These identifications affect the course and outcome of criminal investigations and convictions. Despite calls for research on

- **Thanks**

- **Dr. Richard Vorder Bruegge**
- **FISWG**

# Two Dimensions of Recognition



# Motivation and Goals

- Motivation
  - Forensic 1-to-1 Comparison
    - Testify in court
  - Meet Daubert criteria (US legal system)
- Goals
  - Performance of forensic examiners
    - Comparison to population

# Perceptual Test



- Human subject raters respond...
  - 1. sure they are the same person
  - 2. think they are the same person
  - 3. not sure
  - 4. think they are not the same person
  - 5. sure they are not the same person

# Questions Asked

- Are forensic examiners better than the general population?
- Does time looking at a face matter?
- Do examiners look at more than the face?
- Do examiners recognize faces differently?

# Basics

- 6 May 2014 Facial Identification Scientific Working Group Meeting, Quantico VA
  - 27 Examiners (international group)
  - 14 Non-examiners (controls)
- UNSW
  - 32 Student volunteers (students)

# Three Tests — Six Tasks

- Glasgow Face Matching Test
- Expertise in Facial Comparison Test
  - 2 second and 30 second exposure time
  - Upright and inverted faces
- Person Identification Test



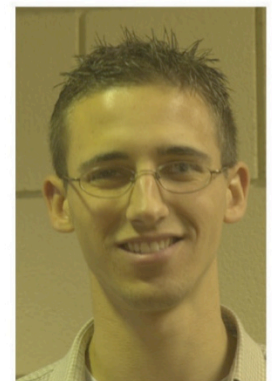
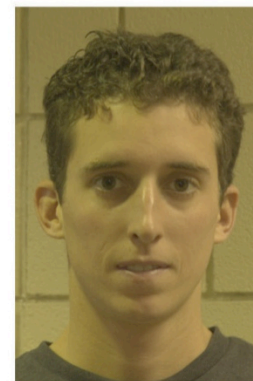
# Expertise in Facial Comparison Test

Question: Does time looking at a face matter?

Match

Non-match

Upright  
2 Sec & 30 Sec



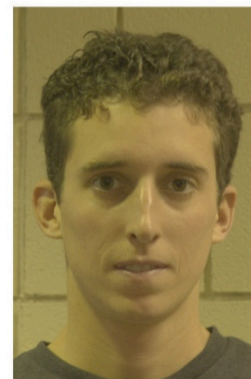
# Expertise in Facial Comparison Test

Question: Do examiners process faces differently?

Match

Non-match

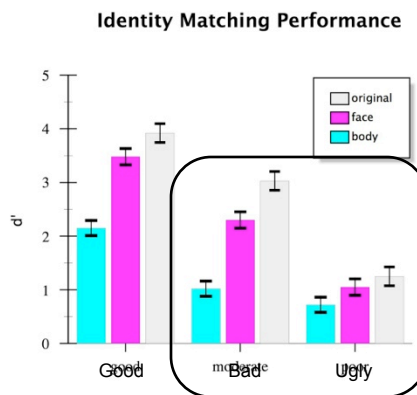
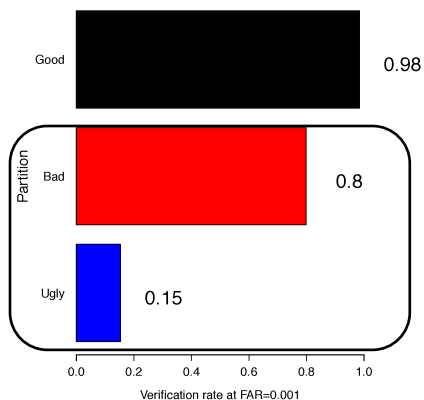
Upright  
2 Sec & 30 Sec



Inverted  
2 Sec & 30 Sec



# Creation of Expertise in Facial Comparison Test

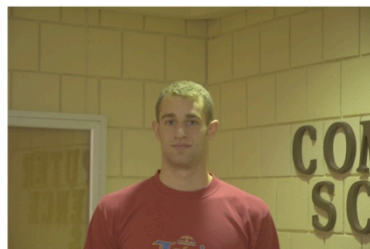


**The Good, Bad, and Ugly Face Challenge**

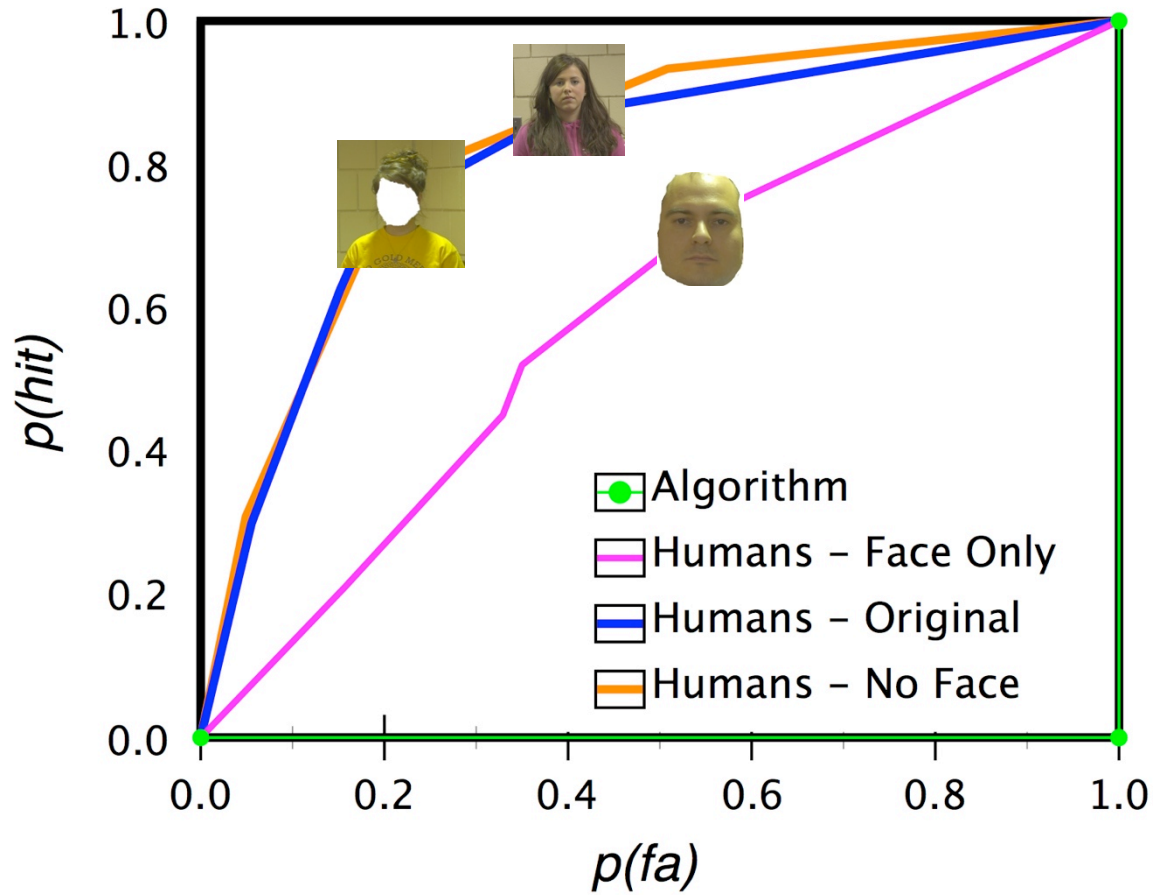


# Person Identification Test

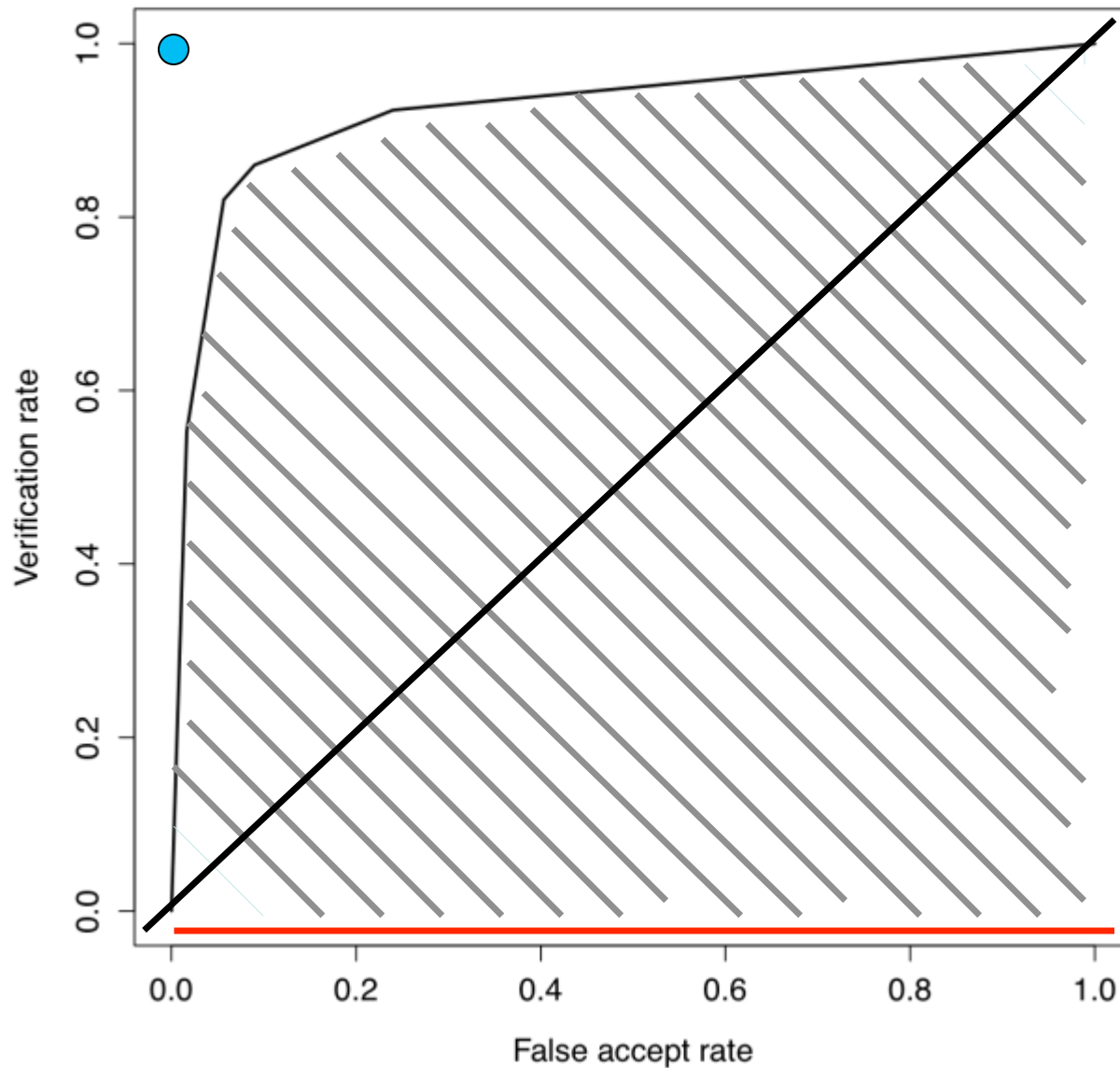
Question: What is the role of the body?



# Human vs. Algorithm (100% Incorrect)



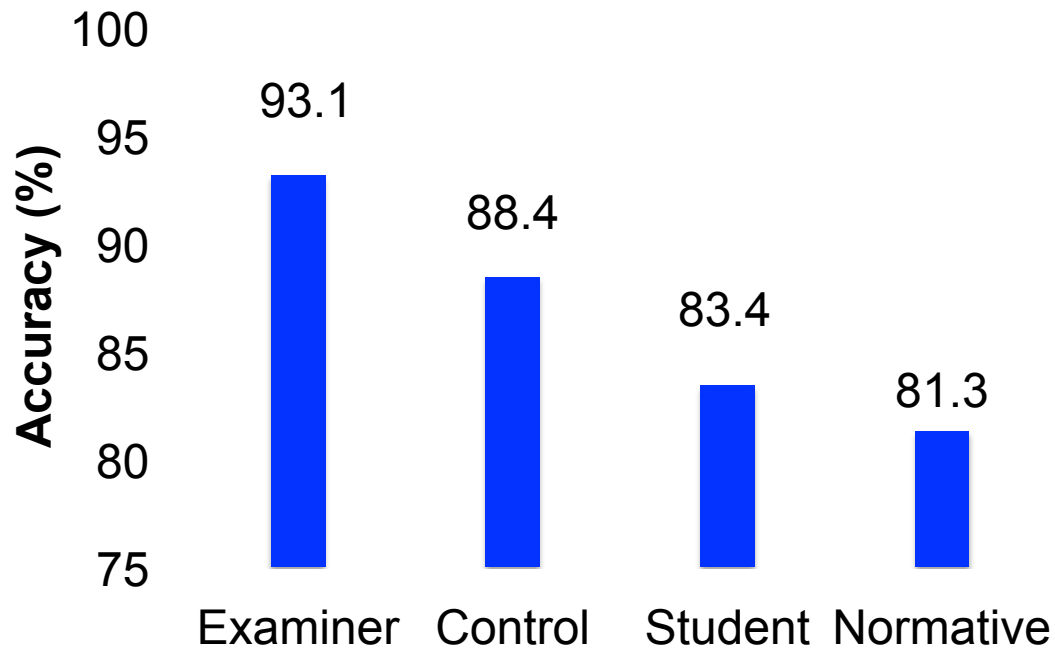
# Area Under Curve (aROC)



# Glasgow Face Matching Test—Results

**Answer: Examiners better than normative (general) population**

First known occurrence of groups better than normative



- Examiner > Normative  $t(219) = 6.35; p = 0.00001$
- Control > Normative  $t(206) = 2.77; p = 0.006$



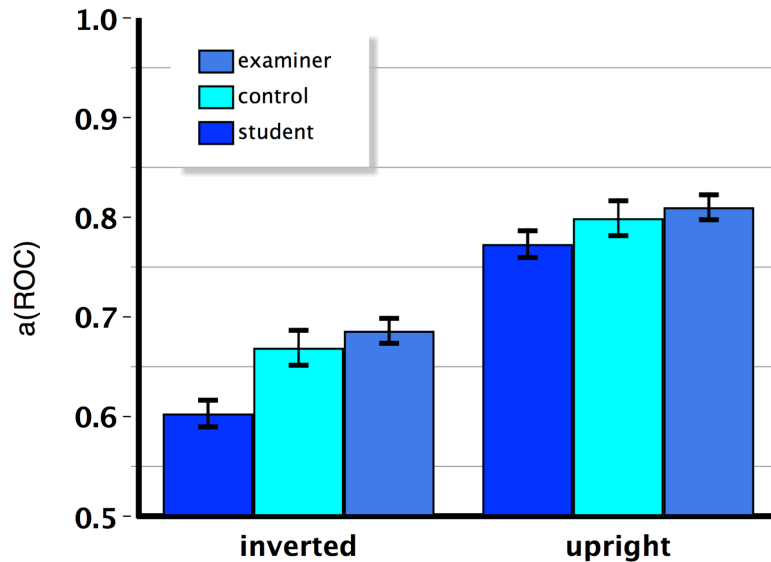
# Expertise in Facial Comparison Test— Results with aROC

Answers: 30 seconds better than 2 seconds

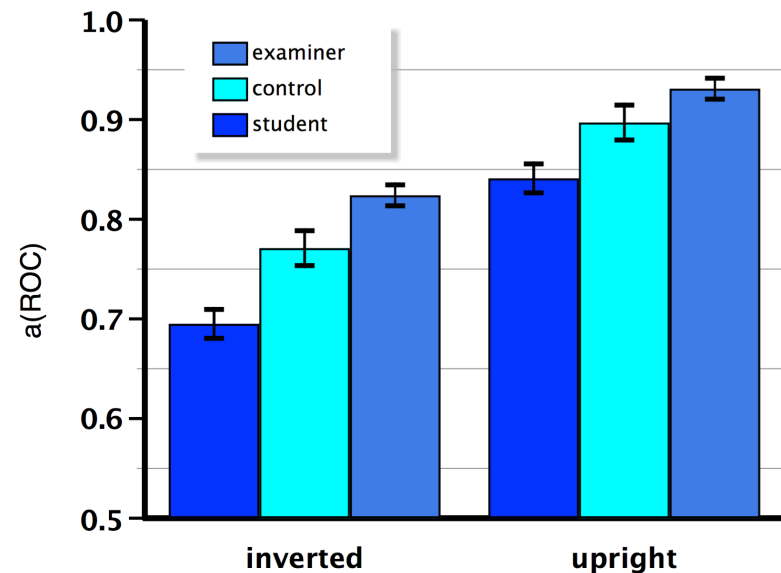
Order of Examiners, Controls, Students

Upright better than Inverted

Expertise in Facial Comparison Test:  
2 seconds



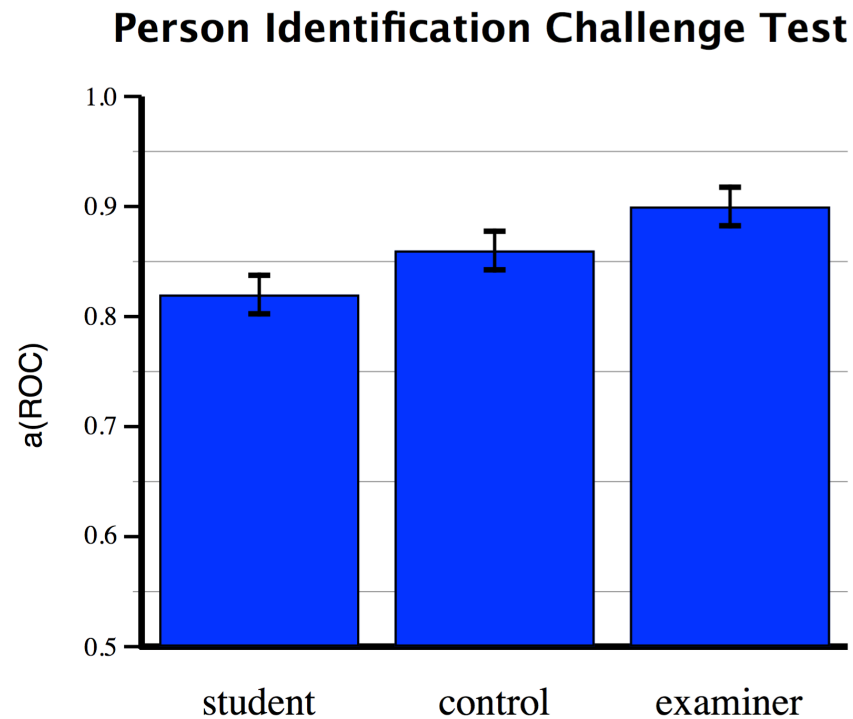
Expertise in Facial Comparison Test:  
30 seconds



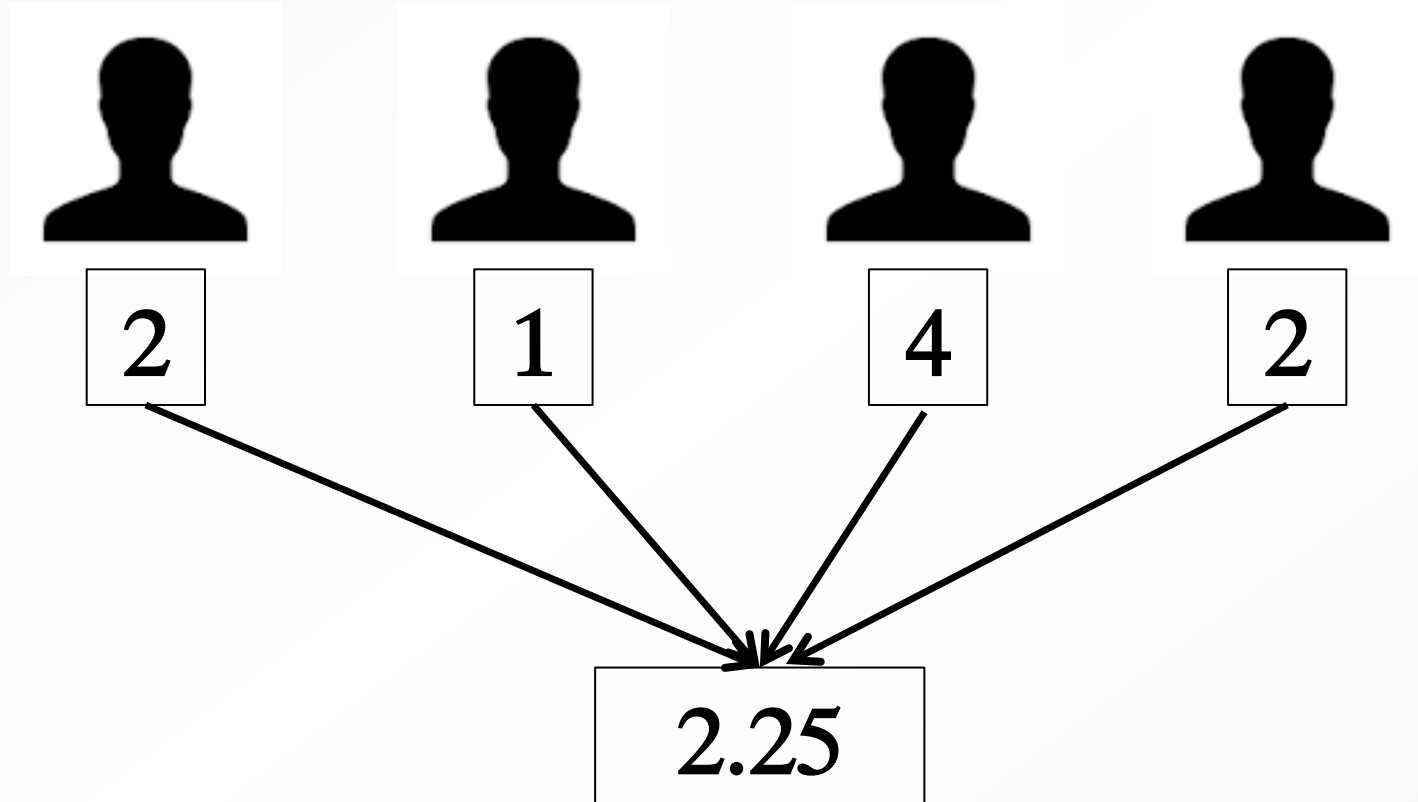
# Person Identification Test—Results

Answers: Examiners appear to use all information

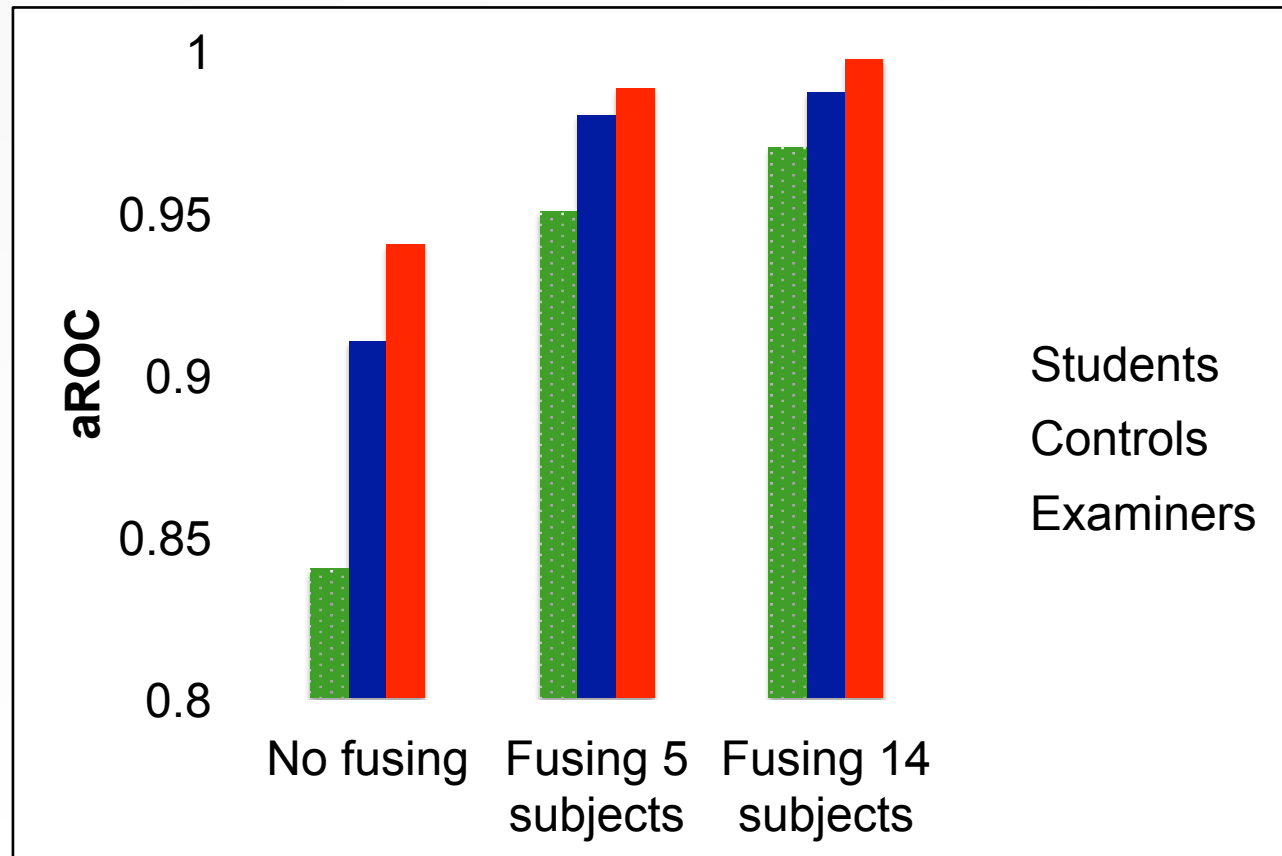
Order of Examiners, Controls, Students



# Fusing Human Ratings



# Fusing Human Ratings



# Overall Results

- On all six tasks, ordering of performance by area under the ROC

**Examiners > Controls > Students**

- Statistical inference
  - Examiners > Controls
  - Controls > Students
  - Wilcoxon sign test, ( $t(5) = 4.85$ ,  $p\text{-value} = .0313$ )

# Conclusions for Perceptual Study

- Examiners out perform general population
- Order of accuracy: Examiners, Controls, and Students
- Time matters
  - 30s better than 2s
- Face and person recognition
  - All identity cues
- Fusion is effective
- Experiments suggest that examiners recognize face differently

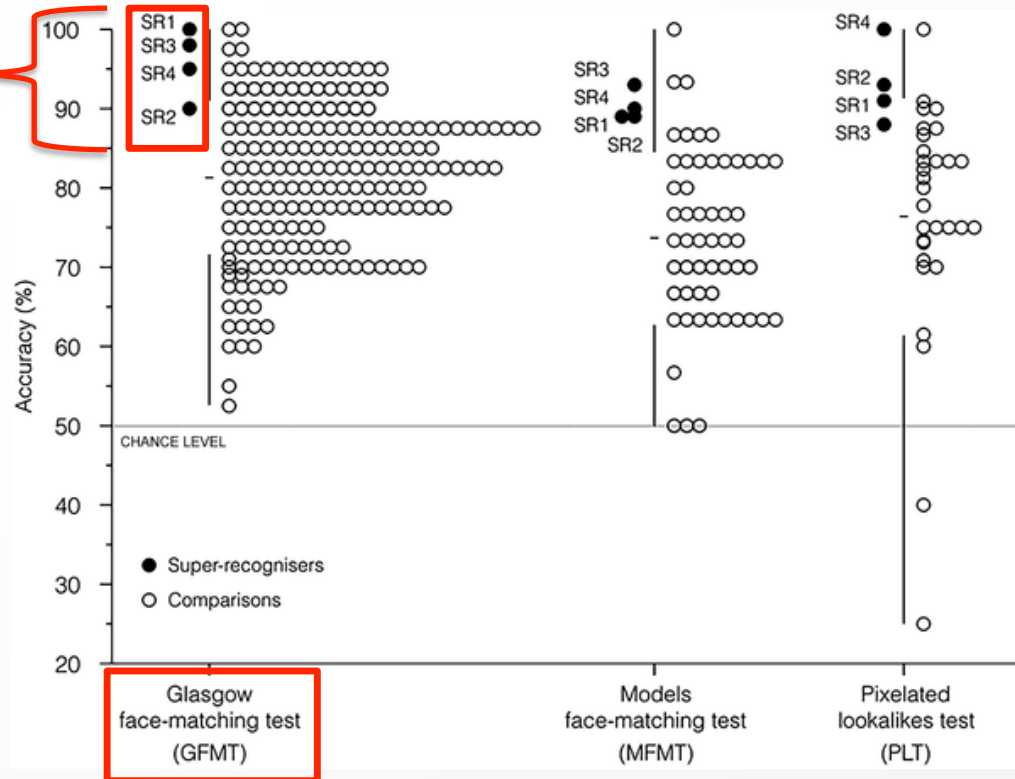
RESEARCH ARTICLE

# Face Recognition by Metropolitan Police Super-Recognisers

David J. Robertson, Eilidh Noyes, Andrew J. Dowsett, Rob Jenkins, A. Mike Burton\*

Fig 2. Performance of police super-recognisers and comparison viewers.

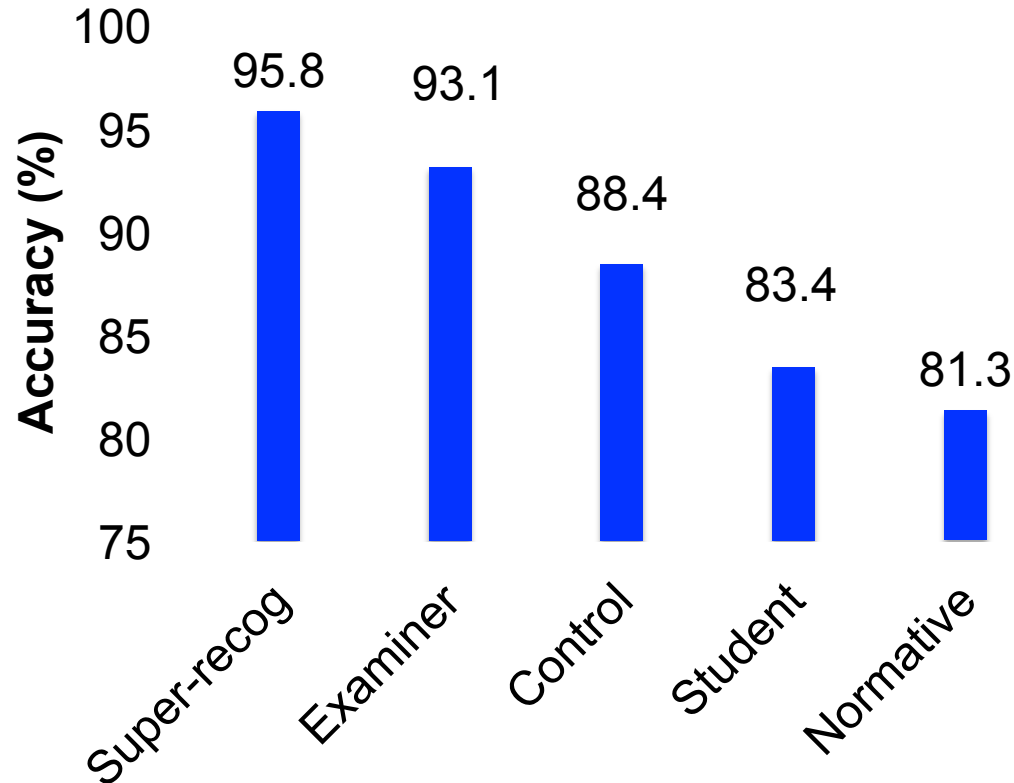
Mean accuracy of super-recognizers 95.8% on GFMT





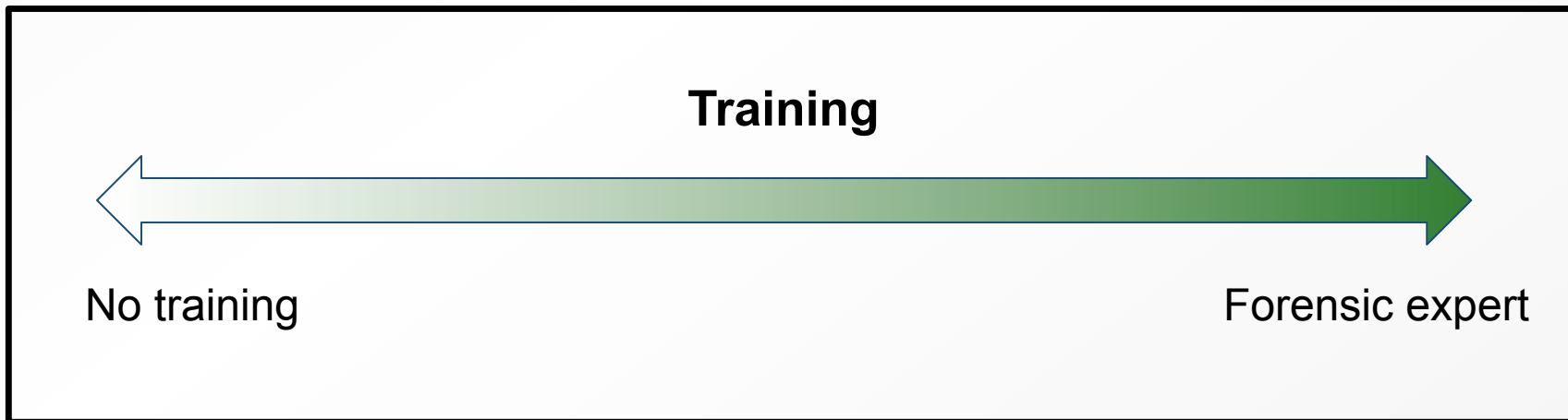
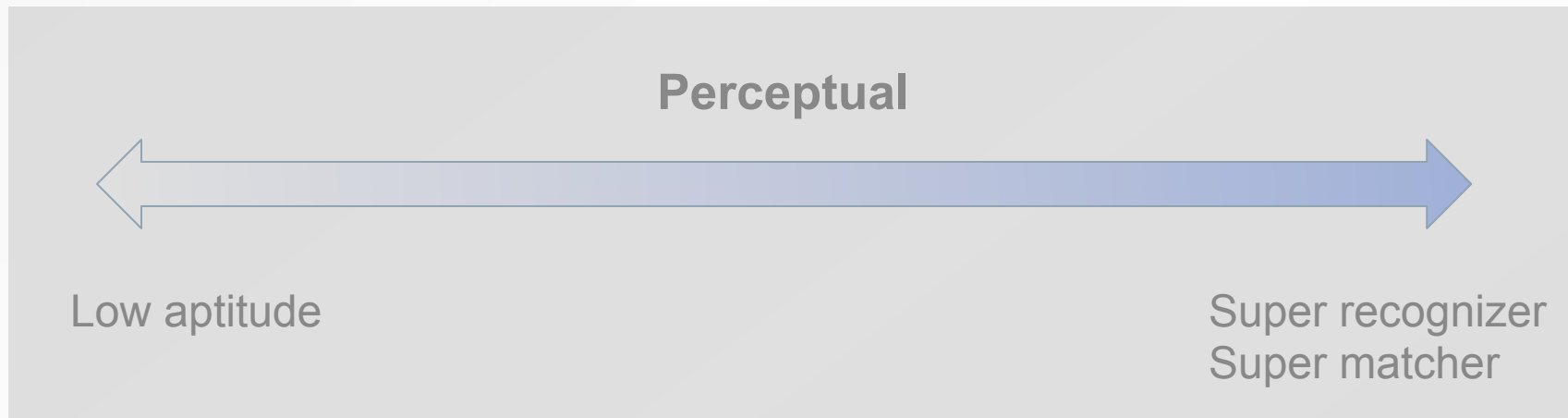
# Glasgow Face Matching Test—Results

Analysis with super-recognizers added



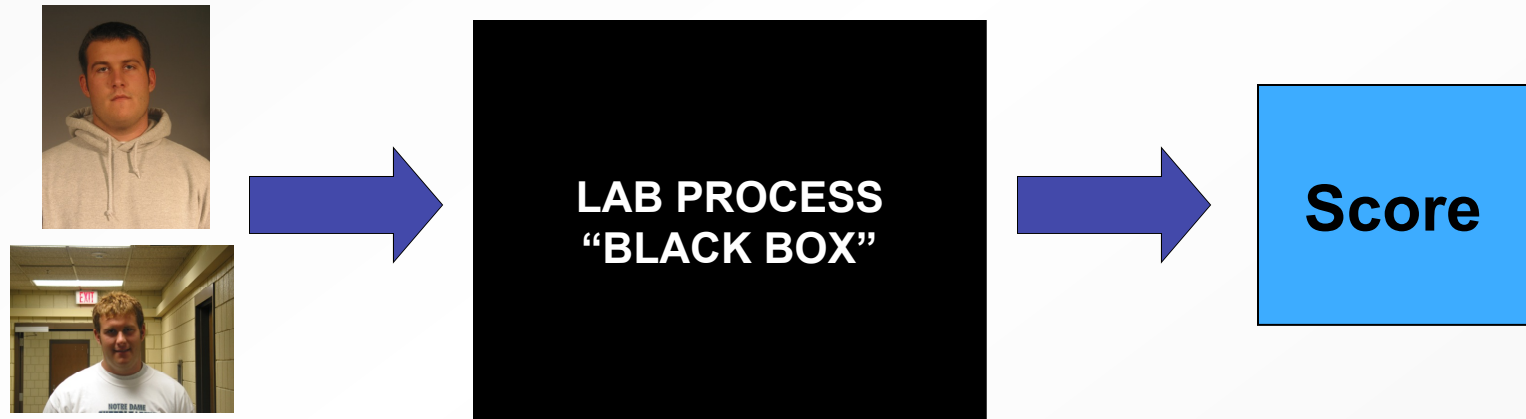
- Examiner > Normative  $t(219) = 6.35; p = 0.00001$
- Control > Normative  $t(206) = 2.77; p = 0.006$

# Two Dimensions of Recognition



# Next Facial Forensic Study

- Measure performance of Forensic Facial Examiners using their tools and process(es).



- Examiners can use lab procedures, tools, methods, resources, and time schedule (more or less).

# The Black-box Team

- NIST
  - Dr. P. Jonathon Phillips
  - Amy Yates
- U of Texas at Dallas
  - Prof. Alice J. O'Toole
- U of New South Wales
  - Dr. David White

# Overview of Black-box Study

- This is an overview
  - Details of the study are in the NIST approved consent form
- Status: **Recruiting**
  - Volunteers from FOUR continents

# General Rules

- Survey questionnaire
- 7 point comparison scale
- 5 point difficulty of comparison scale
- 20 pairs of face images
- 3 months to complete comparisons
- Option to get performance on the test

# Three Subject Groups

- Facial forensic examiners
- Non-examiner face experts
- Fingerprint examiners with no face experience




# Comparison Scale

- +3 The observations strongly support that it is the same person
- +2 The observations support that it is the same person
- +1 The observations support to some extent that it is the same person
- 0 The observations support neither that it is the same person nor that it is different persons
- 1 The observations support to some extent that it is not the same person
- 2 The observations support that it is not the same person
- 3 The observations strongly support that it is not the same person



# Difficulty of Comparison

|                |  |
|----------------|--|
| Easy           | The comparison was easier than most facial comparisons.  |
| Moderate       | The comparison was a typical facial comparison.  |
| Difficult      | The comparison was more difficult than most facial comparisons.  |
| Very difficult | The comparison was unusually difficult, involving significant photometric, illumination, or pose changes, other red flags. |
| Not possible   | The comparison was virtually impossible, due to a lack of detail in the image(s).  |



# How Do I Participate?

- We are Recruiting and Enrolling
- Recruiting email to IBPC attendees
- Interested participants please email me
  - [jonathon@nist.gov](mailto:jonathon@nist.gov)

# Questions